[redacted] **JP6259198** Biblio | Page 1 | Drawing [redacted]

DISK DEVICE SYSTEM

Patent Number: JP6259198
Publication date: 1994-09-16
Inventor(s): MATSUMOTO JUN; others: 03
Applicant(s): HITACHI LTD
Requested Patent: JP6259198
Application Number: JP19930048051 19930309
Priority Number(s):
IPC Classification: G06F3/06; G11B19/02; G11B20/10; G11B20/12
EC Classification:
Equivalents: JP3250861B2

Abstract

PURPOSE:To attain the high speed of a command processing in disk device system. CONSTITUTION:A command scheduler 32 which receives a command from a host computer is equipped with a command waiting queue 33 in which arriving commands wait, command rearrangement queue 34 used for the rearrangement of the commands, queue control part 35 which controls the queue, queue size control means 36 which controls the size of the queue, and execution sequence optimization control part 37. The execution sequence optimization control part 37 is equipped with a continuous area access command rearranging means 38 which rearranges the commands for performing an access to continuous areas, and a multiple execution possibility rearranging means 39 which rearranges the commands to be multiple-executed. The commands transmitted from a host computer are rearranged so that the access to the continuous areas can be performed by the address of data, and so that the multiple execution can be attained.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平6－259198

(43)公開日　平成６年(1994)９月16日

(51)Int.Cl.[5]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ	3/06	３０２　Ｅ	7165－5B		
Ｇ１１Ｂ	19/02	Ｆ	7525－5D		
	20/10	Ａ	7736－5D		
		Ｄ	7736－5D		
	20/12		9295－5D		

審査請求　未請求　請求項の数６　ＯＬ　（全 18 頁）

(21)出願番号　特願平5－48051

(22)出願日　平成５年(1993)３月９日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　松本　純
神奈川県横浜市戸塚区吉田町292番地　株式会社日立製作所マイクロエレクトロニクス機器開発研究所内

(72)発明者　磯野　聡一
神奈川県横浜市戸塚区吉田町292番地　株式会社日立製作所マイクロエレクトロニクス機器開発研究所内

(74)代理人　弁理士　富田　和子

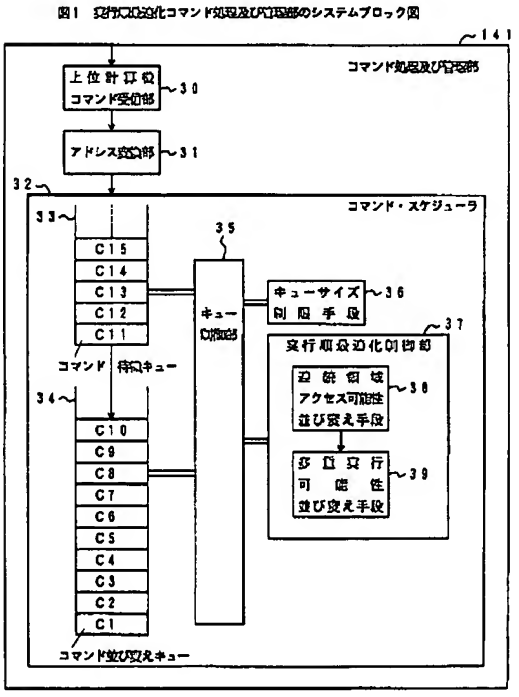
最終頁に続く

(54)【発明の名称】　ディスク装置システム

(57)【要約】

【目的】　ディスク装置システムにおけるコマンド処理を高速化することにある。

【構成】　上位計算機からのコマンドを受けるコマンド・スケジューラ３２は、到着コマンドが待機するコマンド待機キュー３３と、コマンドの並び変えに用いるコマンド並び変えキュー３４と、キューを制御するキュー制御部３５と、キューのサイズを制限するキューサイズ制限手段３６と、実行順最適化制御部３７とを有する。実行順最適化制御部３７は、連続領域をアクセスするコマンドを並び変える、連続領域アクセスコマンド並び変え手段３８と、多重実行可能なコマンドを並び変える多重実行可能性並び変え手段３９とを有し、上位計算機から送出されたコマンドをデータのアドレスにより連続領域をアクセスするように並び変えることと、多重実行が可能なように並び変える。



1

【特許請求の範囲】

【請求項1】ディスク装置と上記ディスク装置を制御するディスク制御装置とを有し、外部からの、データの読み出しまたは書き込みのコマンドを受けて、上記データの読み出しまたは書き込みを行うディスク装置システムにおいて、

上記ディスク制御装置は、

上記コマンドを受信する受信手段と、

上記コマンドを蓄積する第一コマンドキューと、

上記蓄積されたコマンドの実行時間の総和を短縮するように上記コマンドの実行順序を変更する実行順制御手段とを有することを特徴とするディスク装置システム。

【請求項2】請求項1記載のディスク装置システムにおいて、

上記実行順制御手段は、上記コマンドが処理対象とするデータのアドレスを調べ、アドレスが連続するコマンドを抽出し、上記アドレスが連続するように、上記コマンドの実行順序を変更することを特徴とするディスク装置システム。

【請求項3】請求項1記載のディスク装置システムにおいて、

上記ディスク装置システムは、複数のディスク装置を有し、

上記実行順制御手段は、上記コマンドが処理対象とするディスク装置を調べ、異なるディスク装置を処理対象とするコマンドが連続するように、上記コマンドの実行順序を変更し、

異なるディスク装置が並行して処理をすることを特徴とするディスク装置システム。

【請求項4】請求項1記載のディスク装置システムにおいて、

上記実行順制御手段は、

上記コマンドが処理対象とするデータのアドレスを調べ、アドレスが連続するコマンドを抽出し、上記アドレスが連続するように、上記コマンドの実行順序を変更し、

次に、上記連続するアドレスを有する複数のコマンドは1つのコマンドとみなして、上記コマンドが処理対象とするディスク装置を調べ、異なるディスク装置を処理対象とするコマンドが連続するように、上記コマンドの実行順序を変更することを特徴とするディスク装置システム。

【請求項5】請求項1、2、3または4記載のディスク装置システムにおいて、

予め定められた条件を満たしたときに、上記第一コマンドキューにコマンドを蓄積することを停止するキューサイズ制限手段と、

制限されて、上記第1コマンドキューに蓄積されなくなったコマンドを蓄積する第二コマンドキューと、

第一コマンドキューに蓄積したコマンドの並び変えが終了後、上記コマンドを送出し、次に、第二コマンドキューにあるコマンドを第一コマンドキューに送るキュー制御手段とを有することを特徴とするディスク装置システム。

2

【請求項6】請求項1記載のディスク装置システムにおいて、

上記キューサイズ制限手段は、第一コマンドキューにコマンドを蓄積することを、コマンド数が一定値に達した場合、または一定時間が経過した場合、または特定のコマンドを外部から受信した場合に停止することを特徴とするディスク装置システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、ハード・ディスク、光ディスク及び半導体ディスク等のディスク装置を使用するディスク装置システムにおける、大容量データの高速書き込み、読み出し処理に関する。複数ディスク装置で構成されるディスクアレイのコマンド処理にも適する。

【0002】

【従来の技術】ファイルシステムの大容量化とトランザクション性能の向上が求められる中、大量のデータを高速に処理する一方式として特開平2－236714号公報に開示されるアレイ型ディスク駆動機構システム及び方法がある。この方式を図10を用いて説明する。

【0003】図10において、10はRead／Write コマンドを発行する上位計算機、11はデータの分配・集合制御を行うアレイ・コントローラである。11の中で、12は上位計算機とのコマンド及びデータの授受を行う上位計算機I／F部、13はタスク実行管理を行うタスク制御部、14はコマンド多重処理の最適化を行うコマンド処理管理部、及び上位計算機12からのコマンドをディスクアレイのコマンドに再編成するアドレス変換や各ディスク装置へのコマンド生成、発行、終了処理を行うコマンド処理部、15は割込み処理部、16～20はディスクとのコマンド及びデータ授受を行うディスク装置I／F部、21～25はデータの書き込み、読みだしを行うディスク装置である。なお、ここで扱うコマンドはSCSIコマンドであり、それ自体がコマンドの種類、論理ブロックアドレス（以下LBAと記す）、転送長の3つの要素を持っている。

【0004】まず、図10に示すアレイ・コントローラ11によって複数のドライブ装置21～25を多重動作（並行に動作させること）させる。上位計算機10から転送されるコマンド・データは、アレイ・コントローラ11のコマンド処理管理部15にて、各ディスク装置に所定の単位で振り分けるためにアドレス変換される。各コマンドをファーストイン・ファーストアウト方式（以下FIFO方式と記す）により到着順にタスク制御管理部13や割込み処理部15の管理の下で、ディスク装置I／F部16～20を介して、各ディスク装置21～2

5に書き込む。コマンド・データの読み出しの場合には、各ディスク装置21～25から読み出された分割されたコマンド・データをアレイ・コントローラ11で1つのデータに統合し、上位計算機10に送出する。

【0005】

【発明が解決しようとする課題】かかる従来の方法においては以下の課題がある。

【0006】すなわち、従来のコマンドを到着順に処理するFIFO方式では、上位計算機が複数ディスクに対してRead／Writeコマンドを実行する場合、連続データをアクセスする複数のコマンド列中に、該連続データに無関係なデータをアクセスするコマンドが混在するような場合、その混入毎にヘッドシークが発生し、著しく連続データ転送特性を低下させるという課題がある。

【0007】また、あるコマンドがディスク装置0とディスク装置1の2台のディスクを使用するような場合、ディスク装置0がアクセス待ちをしていると、ディスク装置1は未使用にもかかわらず、そのコマンドに確保されたままになっているために、後から到着するコマンドがディスク装置1を利用することができず、多重動作が制限されるという課題がある。

【0008】本発明の目的は、コマンドの実行順序を変更して、処理の高速化を実現したディスク装置システムを提供することである。

【0009】

【課題を解決するための手段】上記の目的を達成するため、ディスク装置と上記ディスク装置を制御するディスク制御装置とを有し、外部からの、データの読み出しまたは書き込みのコマンドを受けて、上記データの読み出しまたは書き込みを行うディスク装置システムにおいて、上記ディスク制御装置は、上記コマンドを受信する受信手段と、上記コマンドを蓄積する第一コマンドキューと、上記蓄積されたコマンドの実行時間の総和を短縮するように上記コマンドの実行順序を変更する実行順制御手段とを有することとしたものである。

【0010】

【作用】コマンドが第一コマンドキューに入り、上記蓄積されたコマンドの実行時間の総和を短縮するように上記コマンドの実行順序を変更することにより、高速化が図れる。

【0011】たとえば、実行順制御手段が第一コマンドキューのコマンド群について、連続領域アクセス可能性を考慮して並び変え、並び変えが終わったコマンドから実行を開始することで、連続領域アクセスデータをまとめて処理することができる。

【0012】これにより、分断された連続領域アクセスデータをまとめて1つのコマンド群として処理することで、ディスクヘッドのシーク等を減らすことができ、高速化を実現することである。

【0013】また、第一コマンドキューのコマンド群について多重実行可能性を考慮して、並び変え、並び変えが終わったコマンドから実行を開始することで、コマンドを多重実行することができる。

【0014】これにより、複数のディスク装置が極力多重実行可能なように、コマンドを並び変えられるので、ディスクの稼働率を上げ、1コマンドあたりの平均処理時間を短縮することができる。

【0015】

【実施例】本発明の第一実施例を図1～4、11を用いて説明する。

【0016】図11に示す本ディスク装置システム（ディスクアレイ）は、Read／Writeコマンドを発行する上位計算機10から、コマンドを受けてデータの分配・集合制御を行うアレイ・コントローラ11（ディスク制御装置）と、データの書き込み、読みだしを行うディスク装置21～25とを有する。アレイ・コントローラ11は、上位計算機とのコマンド及びデータの授受を行う上位計算機I／F部12と、タスク実行管理を行うタスク制御部13と、コマンド多重処理の最適化を行うコマンド処理並びに管理、及び上位計算機10からのコマンドをディスクアレイのコマンドに再編成するアドレス変換や各ディスク装置へのコマンド生成、発行、終了処理を行うコマンド処理及び管理部141と、割込み処理部15と、ディスクとのコマンド及びデータ授受を行うディスク装置I／F部16～20とを有する。なお、ここで扱うコマンドはSCSIコマンドであり、それ自体がコマンドの種類、論理ブロックアドレス（以下LBAと記す）、転送長の3つの要素を持っている。

【0017】まず、図11に示すアレイ・コントローラ11によって複数のドライブ装置21～25を多重動作（並行に動作させること）させる。上位計算機10から転送されるコマンド・データは、アレイ・コントローラ11のコマンド処理管理部15にて、各ディスク装置に所定の単位で振り分けるためにアドレス変換される。各コマンドをファーストイン・ファーストアウト方式（以下FIFO方式と記す）により到着順にタスク制御管理部13や割込み処理部15の管理の下で、ディスク装置I／F部16～20を介して、各ディスク装置21～25に書き込む。コマンド・データの読み出しの場合には、各ディスク装置21～25から読み出された分割されたコマンド・データをアレイ・コントローラ11で1つのデータに統合し、上位計算機10に送出する。

【0018】図1は、コマンド処理及び管理部141の詳細を表わした図である。

【0019】コマンド処理及び管理部141は、上位計算機からのRead／Writeコマンドを受信する上位計算機コマンド受信部30（受信手段）と、Read／Writeコマンドの持つLBAと転送長から、上位計算機が送出したコマンドをアレイディスクのコマンド

5

に再編成するアドレス変換部３１と、Ｒｅａｄ／Ｗｒｉｔｅコマンドの実行順を並び変えるコマンド・スケジューラ３２とを有する。

【００２０】コマンド・スケジューラ３２は、到着コマンドが待機するコマンド待機キュー３３（第２コマンドキュー）と、コマンドの並び変えに用いるコマンド並び変えキュー３４（第１コマンドキュー）と、コマンド待機キュー３３、コマンド並び変えキュー３４を制御するキュー制御部３５（キュー制御手段）と、キューのサイズを制限するキューサイズ制限手段３６と、実行順最適化制御部３７（実行順制御手段）とを有する。

【００２１】コマンドの実行順最適化制御部３７は、連続領域をアクセスするコマンドを並び変える、連続領域アクセスコマンド並び変え手段３８と、多重実行可能なコマンドを並び変える多重実行可能性並び変え手段３９とを有する。

【００２２】図２のフローチャートに基いて、図１の全体動作を説明する。

【００２３】図１に示す上位計算機コマンド受信部３０はＲｅａｄ／Ｗｒｉｔｅコマンドを受信し（図２フローチャート２００）、アドレス変換部３１はアレイディスクにデータを分配するアドレス変換をコマンドに施す（同２０２）。コマンド・スケジューラ３２はコマンドを取り入れ、それを到着順にコマンド待機キュー３３にセットする（同２０４）。コマンドはコマンド待機キュー３３を通過して、コマンド並び変えキュー３４に入る。ある条件を設けることによって、キューサイズ制限手段３６はコマンド待機キューのコマンド通過を停止する。したがって、コマンド並び変えキュー内のコマンド数は限定される（同２０６）。尚この条件には、例えばタイマーによってある一定時間経過後にキューを制限するタイミングを送出する方法及び、コマンド並び変えキューに到着するコマンド数をあらかじめ限定しておく方法がある。また、これから到着するコマンドは、コマンド並び変えキュー３４で前に到着したコマンドの並び変えが終了するまで、コマンド待機キュー３３で待機する。コマンド並び変えキュー内のコマンドについては到着順を前提に（同２０８）、初めに連続領域アクセス可能性（同２１０）、次に多重実行可能性（同２１２）の有無をチェックし、両方存在する場合は連続領域アクセス可能性を優先的に考慮してコマンドを並び変え、連続領域をアクセスしないコマンドについてのみ多重実行可能性を考慮してコマンドを並び変える。またどちらか一方が存在する場合は、独立してコマンドの並び変えを実施する（同２１４）。また、並び変えが終わったコマンドから実行を開始する（同２１６）。コマンド並び変えキューが空になったら（同２１８）、コマンド待機キュー３３のコマンド通過停止を解除し、待機していた新しいコマンドをコマンド並び変えキュー３４に入れ、以下これまでの動作が繰り返される。

6

【００２４】第一実施例として多重実行可能性のみを考慮した場合を説明する。

【００２５】図３は図１のコマンド・スケジューラ３２に当たる。図３において、３３は到着コマンドが待機するコマンド待機キュー、３４はコマンドの並び変えに用いるコマンド並び変えキュー、４０は３３、３４のキューを制御するキュー制御部、４１はキューのサイズを制限するタイミングを決定するタイマー、４２はアドレスのビット・パターン作成部である。４３は先頭コマンド及び並び変えが終了したコマンドのビット・パターンを読み取る第一ビット・パターン読み取り部、４４はまだ並び変えていないコマンドのビット・パターンを読み取る第二ビット・パターン読み取り部、４５は４３、４４で読み取ったビット・パターンを比較する比較器である。

【００２６】図４のフローチャートに基いて、図３全体の動作を説明する。

【００２７】図３に示すコマンド・スケジューラはアドレス変換されたＲｅａｄ／Ｗｒｉｔｅコマンドを取り入れ（図４のフローチャート３００、３０２）、それを到着順にコマンド待機キュー３３に入れる。この時、ビット・パターン作成部４２はコマンド固有のＬＢＡを元に当該コマンドが使用するディスク装置を示すビット・パターンを作成する。ここでビット・パターンは使用するディスク装置を１、使用しないディスクを０で表わし、図１０のディスク装置の左からそれぞれ対応させたビット列である。例えば、０から２番のディスク装置を使用し、３、４番のディスク装置を使用しない場合、ビット・パターンは１１１００となる。また、ビット・パターン作成部４２は作成したビット・パターンをコマンドに対応づけて保持する（同３０４）。そこで、第一ビット・パターン読み取り部４３はキューの先頭コマンドのビット・パターンを読み取り、反転する（同３０８）。反転したビット・パターンに１がある場合、つまり空いているディスク装置がある場合は、第二ビット・パターン読み取り部４４が、並び変える残りのコマンドのビット・パターンを読み取り、比較器４５がそれを第一ビット・パターン読み取り部のビット・パターンとビット毎に比較し（同３１０）、一致する１の最も多いコマンドを選択する。その後キュー制御部４０はそのコマンドを先頭コマンドの後に並び変える（同３１２）。また、並び変えたコマンドから実行が開始され（同３１４）、コマンド並び変えキューが空になったら（同３１６）、コマンド待機キュー３３のコマンド通過停止を解除し、待機していた新しいコマンドをコマンド並び変えキュー３４に入れ、以下これまでの動作が繰り返される。ビット・パターンの使用はディスクアクセスの多重度を上げるために、使用するディスク装置ができるだけ重ならないようにコマンドを選択する並び変え方式において、簡便で有効な手段である。

7

【００２８】具体例として、図３に示すような１０個のＲｅａｄ／Ｗｒｉｔｅコマンドがコマンド並び変えキュー３４に到着している場合のコマンド並び変え動作を示す。ここでＣ＃はコマンドの到着番号を、（＃＃＃＃＃）は各コマンドのビット・パターンを表わしている。まず、第一ビット・パターン読み取り部４３はＣ１のビット・パターン１１１００（ディスク０、１、２番を使用する）を読み取り、この情報を反転して００００１１を得る。次に第二ビット・パターン読み取り部４４がＣ２からＣ１０までのビット・パターンを読み取り、比較器４５がこの反転ビット・パターンと一致もしくは一致する１の多い（０は全て対応していなければならない）コマンドを検索する。この例では対応するコマンドがないので、Ｃ２が２番目のコマンドになる。こんどは第一ビット・パターン読み取り部４３がＣ２のビット・パターン００１０１を読み取り、このビット・パターンを反転して１１０１０を得る。ここで、Ｃ１の時と同様に、第二ビット・パターン読み取り部４４がＣ３からＣ１０までのビット・パターンを読み取り、比較器４５が対応するコマンドを検索する。すると、１１０００のＣ９が見つかり、実行順を入れかえ、Ｃ２の次に配置する。このＣ２とＣ９は多重実行である。さらに多重実行できる可能性を判断するため、キュー制御部４０がＣ２とＣ９のＥＯＲ（両方０のビットをマスクする）００００１０を作り、第一ビット・パターン読み取り部４３がこれを読み取る。また、第二ビット・パターン読み取り部４２がＣ３からＣ８、Ｃ１０のビット・パターンを読み取り、比較器４３が対応するコマンドを検索するが、ないのでキュー制御部３５がＣ３を４番目のコマンドに並び変える。以下この操作が繰り返される。並び変えた結果を図９（ａ）に示す。

【００２９】この具体例によれば、各コマンド処理を極力多重動作で行うため、たとえば上記例の場合、コマンド処理時間を全てのコマンドで等しく１と仮定すると、実行順を変える前と変えた後では、１０コマンドの処理時間合計は本発明未使用時の１０が、本発明使用により６に削減され、４０％の改善効果がある。したがって、この多重実行最適化によって、コマンドのＩ／Ｏスループットを上げることができる。

【００３０】また、キューを２つ設けることで、並び変えたコマンドが全て実行されてから新しいコマンドの並び変えが行われるため、キューを１つしか用いなかった従来の場合に見られたコマンドの沈み込み（例えば、優先度が低いコマンドは、いつまでもキューの中に実行されないで残る現象）を解消することができる。

【００３１】第二実施例として連続領域アクセス可能性のみを考慮した場合を説明する。

【００３２】図５のコマンドスケジューラ５３２は図１のコマンド・スケジューラ３２に相当する。図５において、３３は到着コマンドが待機するコマンド待機キュー、３４はコマンドの並び変えに用いるコマンド並び変えキュー、５０は３３、３４のキューを制御するキュー制御部、４１はキューのサイズを制限するタイミングを決定するタイマー、５１は連続領域判定アドレス作成手段である。５３７は、実行順最適化制御部である。５１の中で５２は先頭コマンド及びまだ並び変えが終了していないコマンドの内、コマンド・ナンバーの若いもののアドレスを読み取る第一アドレス読み取り部、５３は５２に入るコマンド転送長を読み取る転送長読み取り部、５４は第一アドレス読み取り部５２で読み取ったコマンドの先頭アドレスに転送長を加えて、連続する領域のアドレスを算出する加算機、５６はまだ並び変えていないコマンドのアドレスを読み取る第二アドレス読み取り部、５７は５１と５６のアドレスを比較する比較器である。

8

【００３３】図６のフローチャートに基いて、図５全体の動作を説明する。

【００３４】図５に示すコマンド・スケジューラはアドレス変換されたＲｅａｄ／Ｗｒｉｔｅコマンドを取り入れ（図６のフローチャート４００、４０２）、それを到着順にコマンド待機キュー３３に入れる。コマンドは最初コマンド待機キューを通過してコマンド並び変えキュー３４に入り、タイマー４１が一定時間経過後、並び変えるキューサイズを制限する（同４０４）。この後到着する新しいコマンドはコマンド待機キューで待機する。次に第一アドレス読み取り部５２はコマンド並び変えキュー先頭のコマンドが持つアドレスを読み取り、転送長読み取り部５３はその転送長を読み取り（同４０６）、読み取ったアドレスに転送長を加算器５４で加えることで、連続する領域のアドレスを算出する（同４０８）そして、第二アドレス読み取り部５５は並び変える残りのコマンドが持つアドレスを読み取り、比較器５６がそれを連続領域判定アドレス作成手段５１と比較し（同４１０）、一致するコマンドを先頭コマンドの後に並び変える（同４１２）。また、実行順が決定したコマンドから実行が開始され（同４１４）、コマンド並び変えキューが空になったら（同４１６）、コマンド待機キュー３３のコマンド通過停止を解除し、待機していた新しいコマンドをコマンド並び変えキューに入れ、以下これまでの動作が繰り返される。

【００３５】具体例として、図５に示すような１０個のＲｅａｄ／Ｗｒｉｔｅコマンドがコマンド並び変えキュー３４に到着している場合のコマンド並び変え動作を示す。Ｃ＃はコマンドの到着番号を、（＃、＃）は左がコマンドの持つアドレス、右が転送長を表わしている。まず、連続領域判定アドレス作成手段５１の中の第一アドレス読み取り部５２がＣ１のアドレス０を読み取り、転送長読み取り部５３がＣ１の転送長１６ブロック（１ブロック＝５１２Ｂ）を読み取り、加算機５４はアドレス０に１６を加えて１６を算出する。次に、第二アドレス

9

読み取り部５５がＣ２からＣ１０までのアドレスを読み取り、比較器５６が連続領域判定アドレス作成手段５１で算出された値１６と等しいアドレスを持つコマンドを検索する。すると、Ｃ３が見つかるので、キュー制御部５０がＣ３をＣ１の後に並び変える。さらに連続領域アクセスできる可能性を判断するため、第一アドレス読み取り部５２がＣ３のアドレス１６を読み取り、転送長読み取り部５３がＣ３の転送長１６ブロックを読み取る。前述と同様な操作でコマンドを検索するとＣ６が見つかる。ここまでで、コマンドはＣ１、Ｃ３、Ｃ６の順に並び変えられる。この後Ｃ６について再び連続領域アクセスの可能性を判断すると、今度は対応するコマンドがないため、Ｃ２がＣ６の後に並び変えられる。以下この操作が繰り返される。並び変えた結果を図９（ｂ）に示す。

【００３６】この具体例によれば、１つのコマンドに対するアクセス回数を１とすると、実行順を変える前では連続領域アクセスが可能なコマンド群はＣ８、Ｃ９のみで、アクセス回数は９であるが、実行順を変えた後では連続領域アクセスが可能なコマンド群として、Ｃ１、Ｃ３、Ｃ６及びＣ５、Ｃ７、Ｃ１０も加わるので、アクセス回数は５である。一般的に、十分たくさんのコマンドを連続化することでディスク列全てにアクセスし、ヘッドシーク時間を短縮することができる。また長いデータに他のコマンドが混入する場合のヘッドシーク時間増加を防止する効果もある。

【００３７】さらに、長いデータをアクセスする場合、ディスク装置の持つプリフェッチ・バッファをきかせることができる。すなわち、アドレスが連続しない短いデータを有するコマンドの場合、コマンドごとにプリフェッチ・バッファの内容を書き換える必要が生じる。書き換えの回数が増えることにより処理時間が延びる。

【００３８】本実施例はディスクアレイのみならず、単体のディスク装置に対しても有効なものである。

【００３９】第三実施例として多重実行可能性のみを考慮した場合を説明する。

【００４０】図７は図１のコマンド・スケジューラ３２に当たる。図７において、６１～７０はｋ台のディスク装置の組合せに対応した各コマンド到着キュー（ｎ台のディスク装置からｋ台のディスク装置を選ぶ全ての組合せの数だけ用意したキュー）、７１～８０はコマンド到着キューがコマンドをコマンド実行キューに送出したことを示すフラグ、８１は並び変えを終了したコマンドが実行を待つコマンド実行キューである。６０はコマンド到着キュー６１～７０及びフラグ７１～８０の制御をするキュー制御部、４２はアドレスのビット・パターン作成部、８２はコマンドの並び変えテーブル参照手段である。

【００４１】次に図８のフローチャートに基いて図７全体の動作を説明する。

10

【００４２】図７に示すコマンド・スケジューラ７３２はアドレス変換されたＲｅａｄ／Ｗｒｉｔｅコマンドを取り入れ（図８のフローチャート５００、５０２）、ビット・パターン作成部４２はコマンド固有のＬＢＡを元に当該コマンドが使用するディスクを示すビット・パターンを作成する。次に、キュー制御部６０でそのビット・パターンに対応した各コマンド到着キュー６１～７０に振り分けられる（同５０４、５０６）。コマンド到着キュー先頭のコマンドが全てそろったら、キュー制御部６０がコマンド・ナンバーの最も若いコマンドを選択し（同５０８）、そのコマンドをコマンド実行キュー８１にセットし、フラグ７１～８０のうちコマンドを送出したキューのフラグを立てる（同５１０）。コマンドを送出したキューと対応するキューをコマンド並び変えテーブル参照手段８２で参照し（同５１２）、対応するコマンドがなければ次に若いコマンド・ナンバーのコマンドをコマンド実行キュー８１に送出し、フラグ７１～８０のうちコマンドを送出したキューのフラグを立てる（同５１０）。対応するコマンドがあれば、そのコマンドをコマンド実行キュー８１に送出し、フラグ７１～８０のうちコマンドを送出したキューのフラグを立てる（同５１４）。以上の操作をコマンド到着キュー６１～７０の全てのフラグ７０～８０が立つまで行う（同５１６）。コマンド実行キュー８１に並び変えたコマンドが到着しだい、コマンドを実行し（同５１８）、コマンド実行キュー８１が空になったら（同５２０）、各コマンド到着キュー６１～７０の全てのフラグ７１～８０を解除する（同５２２）。そして、再びコマンド到着キュー６１～７０の中で、最も若いコマンド・ナンバーのコマンドがコマンド実行キュー８１に送出され、以下これまでの動作が繰り返される。

【００４３】具体例として、図７に示すような２台のディスク装置にアクセスする複数のＲｅａｄ／Ｗｒｉｔｅコマンドがコマンド到着キュー６１～７０に到着している場合のコマンド並び変え動作を示す。ここで、Ｃ＃はコマンドの到着番号を表わしている。７３２は、コマンドスケジューラ、６１～７０は到着コマンドが待機するコマンド待機キュー、６０は６１～７０のキューを制御するキュー制御部、７３７は、実行順最適化制御部である。まず、コマンド到着キュー６１～７０の先頭にコマンドが全てそろった状態で、キュー制御部６０がコマンド・ナンバーの最も若いＣ１をコマンド実行キュー８１にセットし、フラグ７１を立てる。Ｃ１は０番と１番のディスク装置を使用するので、このコマンドと多重実行可能なコマンドをコマンド並び変えテーブル参照手段８２において対応づける（図８のテーブル参照）。すると、３番と４番のディスク装置を使用するコマンドが適切ということから、キュー制御部６０がＣ７をコマンド実行キューに送出し、フラグ８０を立てる。次に、キュー制御部６０がフラグのまだ立っていないコマンド到着

11

キューの内、コマンド・ナンバーの最も若いＣ２をコマンド実行キューに送出し、フラグ７６を立てる。Ｃ７の操作と同様にＣ２と多重実行可能なコマンドを検索するとＣ１０が見つかり、キュー制御部６０がＣ１０をコマンド実行キュー８１に送出し、フラグ７９を立てる。以下この操作が繰り返される。並び変えた結果を図９（ｃ）に示す。

【００４４】この具体例によれば、全てのコマンド処理を多重動作で行うため、コマンド処理時間を全て等しく１と仮定すると、実行順を変える前では１０コマンドの処理時間が、実行順を変えた後では５となる。一般に実行順を変える前は、コマンド処理時間はコマンドの並び方によって５～１０の範囲（完全に多重実行する場合から１コマンドも多重実行しない場合まで）で変化するのが、実行順を変えた後では必ず５となるため、最大５０％の改善効果がある。

【００４５】第四実施例として第二実施例の変形例を説明する。

【００４６】図５のコマンド・スケジューラ３２に連続領域でまとめたコマンド群を１つのコマンドとみなす手段を付加する方法がある。これは図９の（ｂ）からも明らかなように、連続領域アクセス可能なコマンド群を１つのコマンドに再構成することで全体の処理コマンド数を減少し、まとめたコマンドについてはそのデータの大きさに関係なくディスク装置のプリフェッチ・バッファが有効利用できることで、プリフェッチ・ミスヒット（連続していないデータの場合、プリフェッチ・バッファにデータが無いことが多いためミスヒットが生じやすい）が避けられ、さらに、ディスク装置の回転待ち時間も回避したＷｒｉｔｅ処理の高速化が実現できる。

【００４７】また、特にディスクアレイでは、Ｗｒｉｔｅ時特有のパリティー更新に伴うＲｅａｄ処理を削減することができるため（データのアドレスが連続していない場合は、例えば、データが異なるとパリティディスクも異なる場合があるため、データごとにパリティディスクを読む必要が生じる）、Ｗｒｉｔｅ処理の高速化が実現できる。

【００４８】以上の説明より、連続領域アクセス可能性及び多重実行可能性双方を考慮したコマンドの並び変えも、図１に示す構成に第一実施例と第二実施例を用いることも可能である。また、第三実施例と第二実施例を用いることによっても第一実施例と第二実施例を用いる場合と同様な効果がもたらされる。また、第二実施例の代わりに第四実施例を用いることでさらに大きな効果がもたらされる。効果としては、実行順を変える前ではコマンド全てが各々Ｒｅａｄ／Ｗｒｉｔｅアクセスする場合でも実行順を変えることで、連続領域アクセス可能なコマンドと多重実行可能なコマンドに整理されることからコマンド実行時間が短縮される。

【００４９】

12

【発明の効果】本発明によれば、コマンドの実行時間の総和が短縮できる。

【００５０】また、連続領域アクセス可能なコマンドをまとめて実行することで、ディスク装置のヘッドシークと回転待ちを短縮するという効果がある。

【００５１】また、コマンドを多重実行することで、ディスク装置の稼動率を上げ、従来に対し最大５０％のＩ／Ｏスループット改善効果がある。

【００５２】さらに、キューのサイズを制限し、その中でコマンドの並び変えを行い、並び変えた順に実行するため、実行されないコマンドがいつまでもキュー内に残ってしまうコマンドの沈み込みを防止するという効果がある。

【図面の簡単な説明】

【図１】本発明に係るコマンド処理及び管理部のブロック図である。

【図２】本発明に係るフローチャートである。

【図３】コマンド・スケジューラの第一実施例を示すブロック図である。

【図４】本発明に係るフローチャートである。

【図５】コマンド・スケジューラの第二実施例を示すブロック図である。

【図６】本発明に係るフローチャートである。

【図７】コマンド・スケジューラの第三実施例を示すブロック図である。

【図８】本発明に係るフローチャートである。

【図９】各実施例によるコマンドの並び変え結果を示す説明図である。

【図１０】ディスクアレイの基本構成を示すブロック図である。

【図１１】ディスクアレイの基本構成を示すブロック図である。

【符号の説明】

１０… …上位計算機、　１１… …アレイ・コントローラ、１２… …上位計算機Ｉ／Ｆ、　１３… …タスク制御部、１４… …コマンド処理及び管理部、　１５… …割込み処理部、１６～２０… …ディスク装置Ｉ／Ｆ、　２１～２５… …ディスク装置、２９… …実行順最適化コマンド処理及び管理部、３０… …上位計算機コマンド受信部、　３１… …アドレス変換部、３２… …コマンド・スケジューラ、３３… …コマンド待機キュー、３４… …コマンド並び変えキュー、　３５… …キュー制御部、３６… …キューサイズ制限手段、　３７… …実行順最適化制御部、３８… …連続領域アクセス可能性並び変え手段、３９… …多重実行可能性並び変え手段、４０… …キュー制御部、４１… …タイマー、　４２… …ビット・パターン作成部、４３… …第一ビット・パターン読み取り部、４４… …第二ビット・パターン読み取り部、４５… …比較器、　５０… ・

13

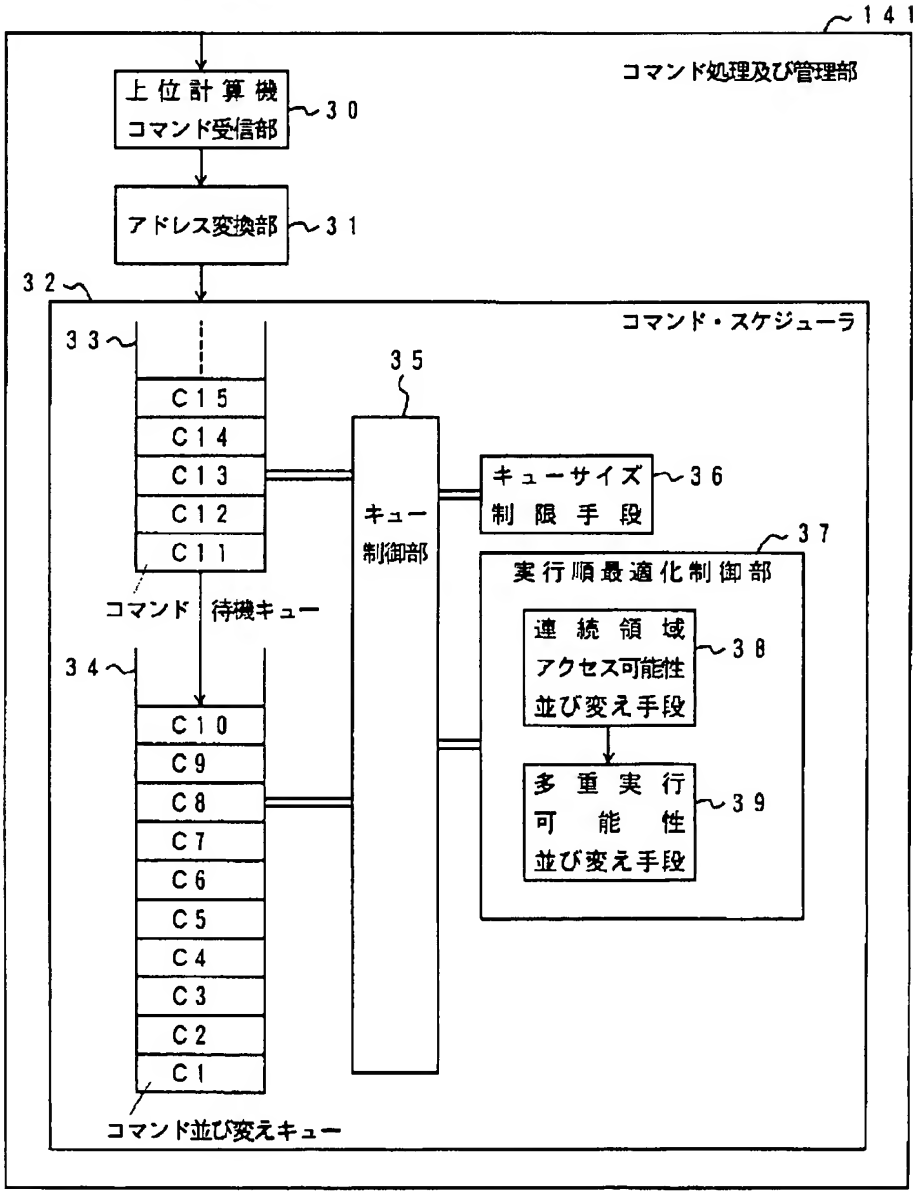
・・キュー制御部、５１・・・ ・・・連続領域判定アドレス作成手段、５２・・・ ・・・第一アドレス読み取り部、　５３・・・ ・・・転送長読み取り部、５４・・・ ・・・加算機、　５５・・・ ・・・第二アドレス読み取り部、５６・・・ ・・・比較器、

14

６０・・・ ・・・キュー制御部、６１～７０・・・ ・・・コマンド到着キュー、　７１～８０・・・ ・・・フラグ、８１・・・ ・・・コマンド実行キュー、８２・・・ ・・・コマンド並び変えテーブル参照手段。

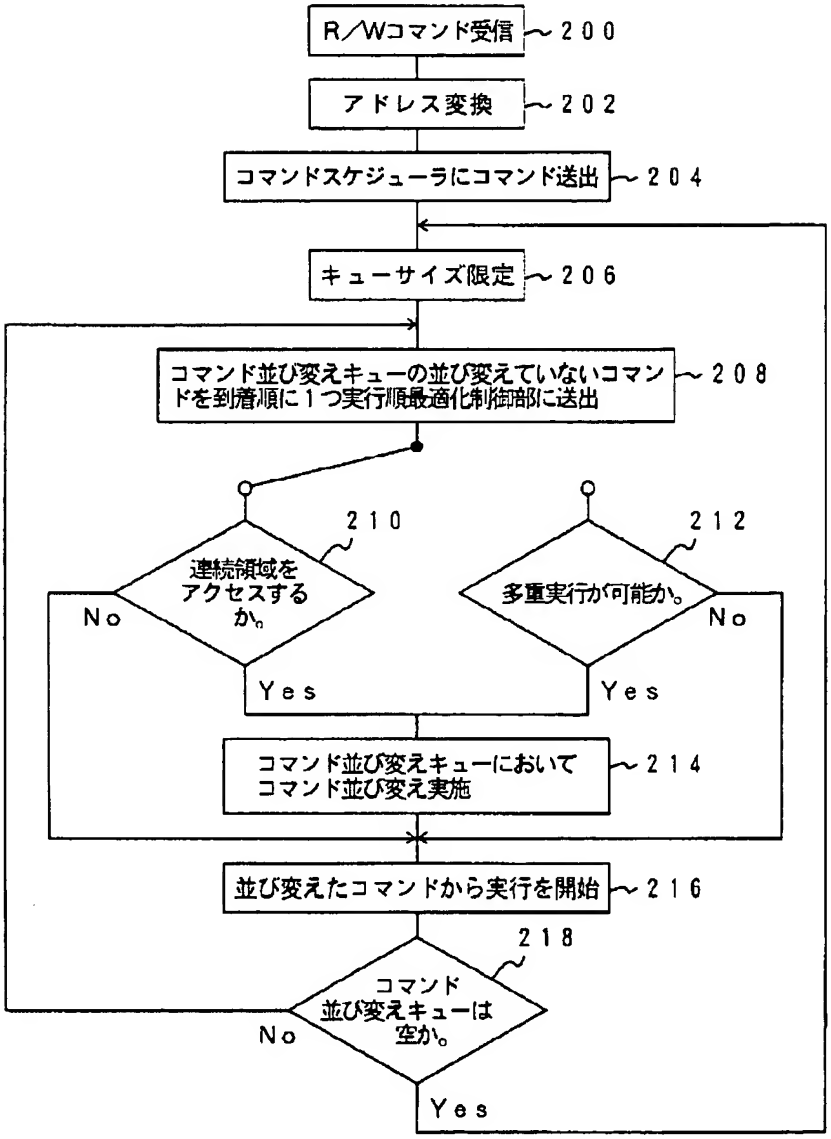
【図1】

図1　実行順最適化コマンド処理及び管理部のシステムブロック図



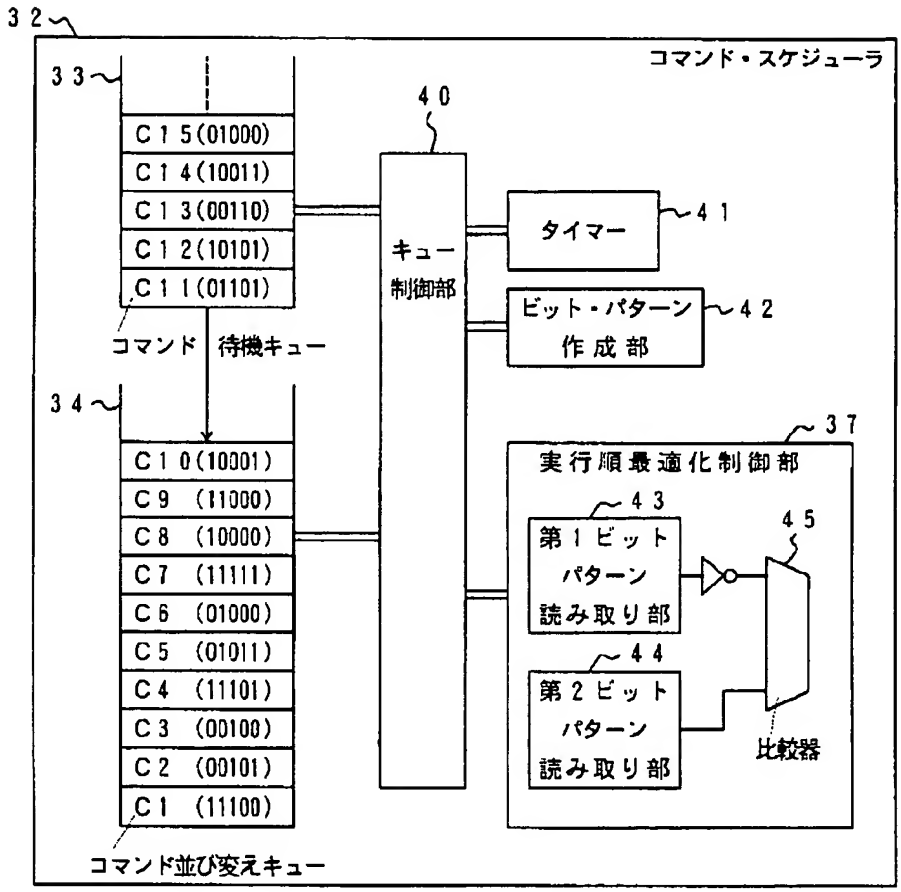
【図2】

図2　本発明の処理手段の基本動作を示すフローチャート



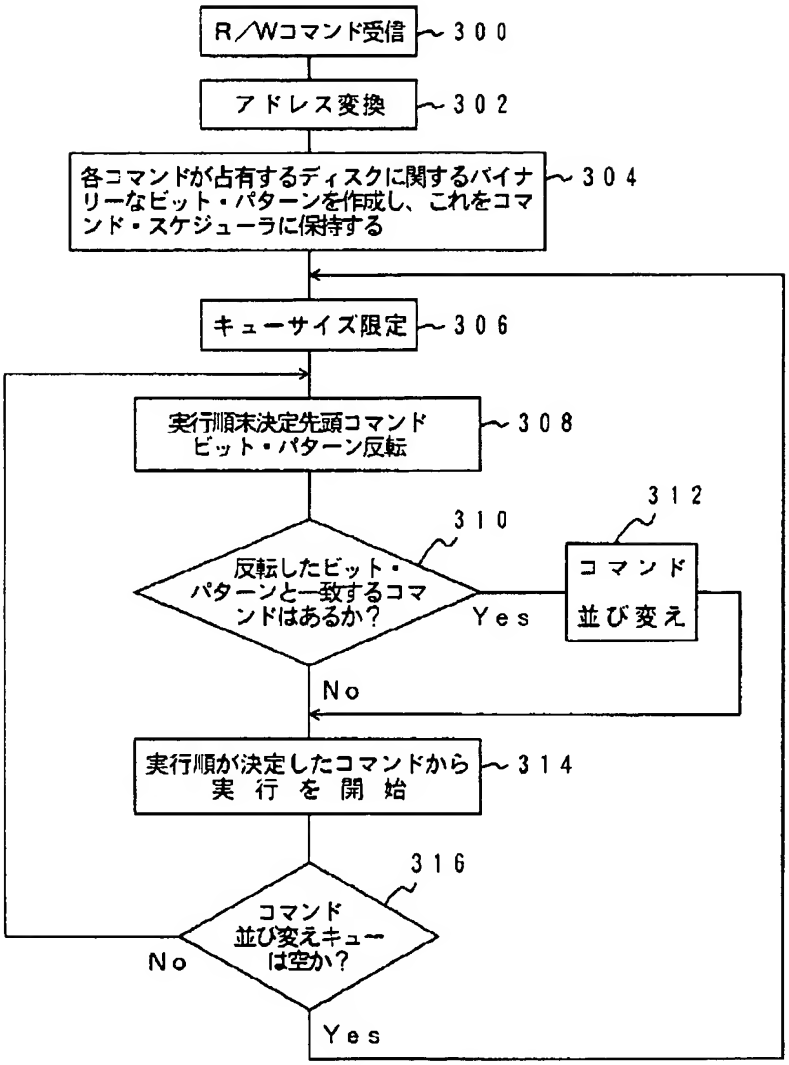
【図3】

図3　図1のコマンド・スケジューラの第一実施例を示すシステムブロック図



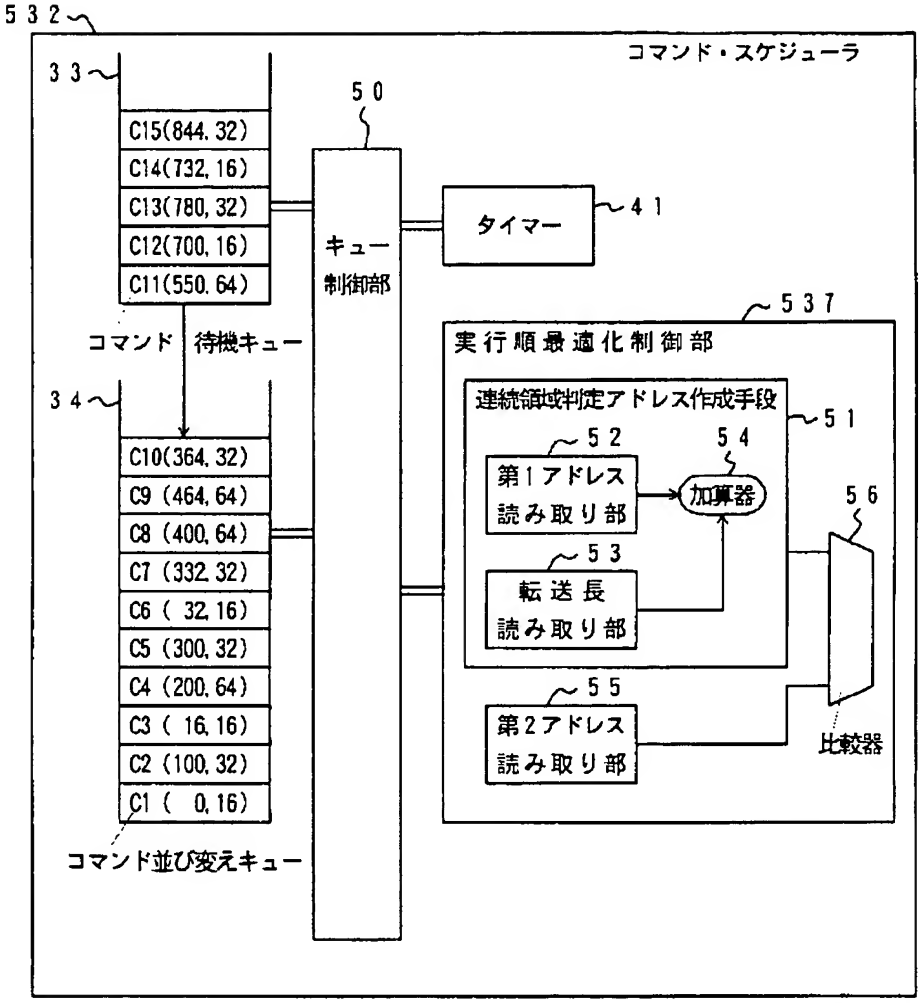
【図4】

図4　第一実施例の全体動作を示すフローチャート



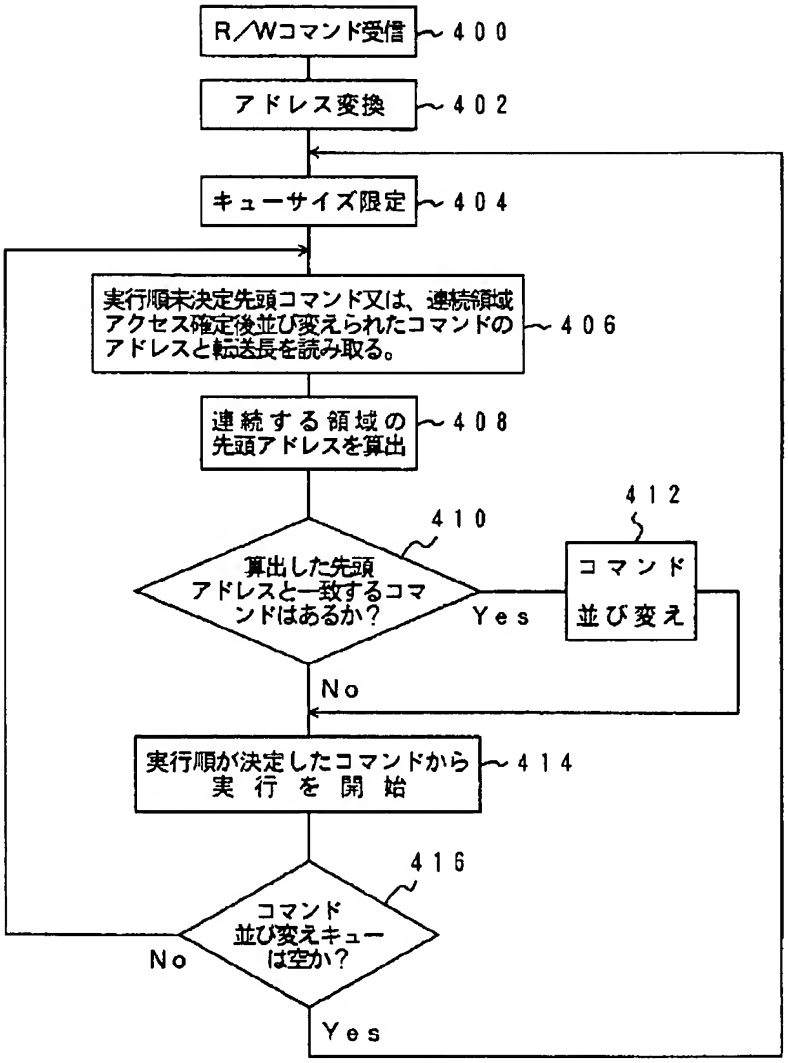
【図5】

図5　コマンド・スケジューラの第二実施例を示すシステムブロック図



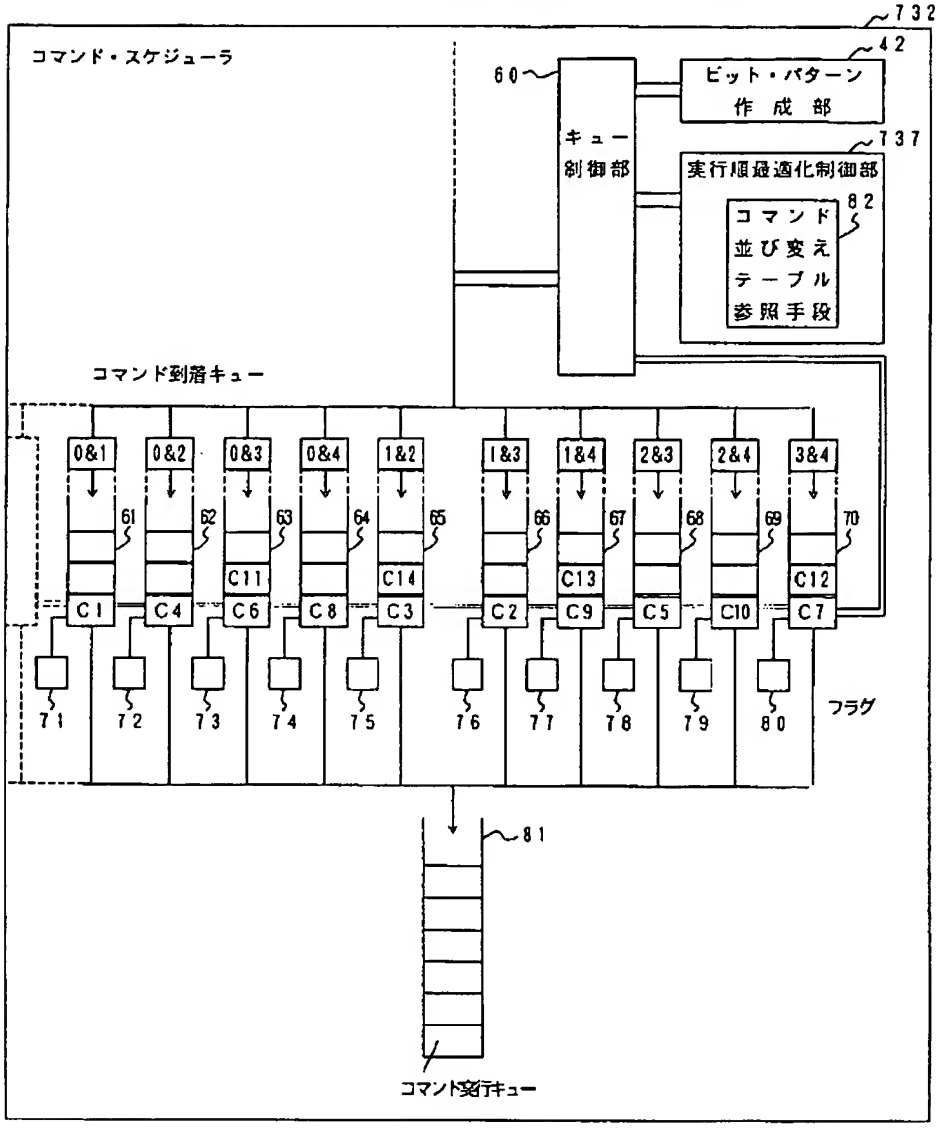
【図６】

図６　第二実施例の全体動作を示すフローチャート



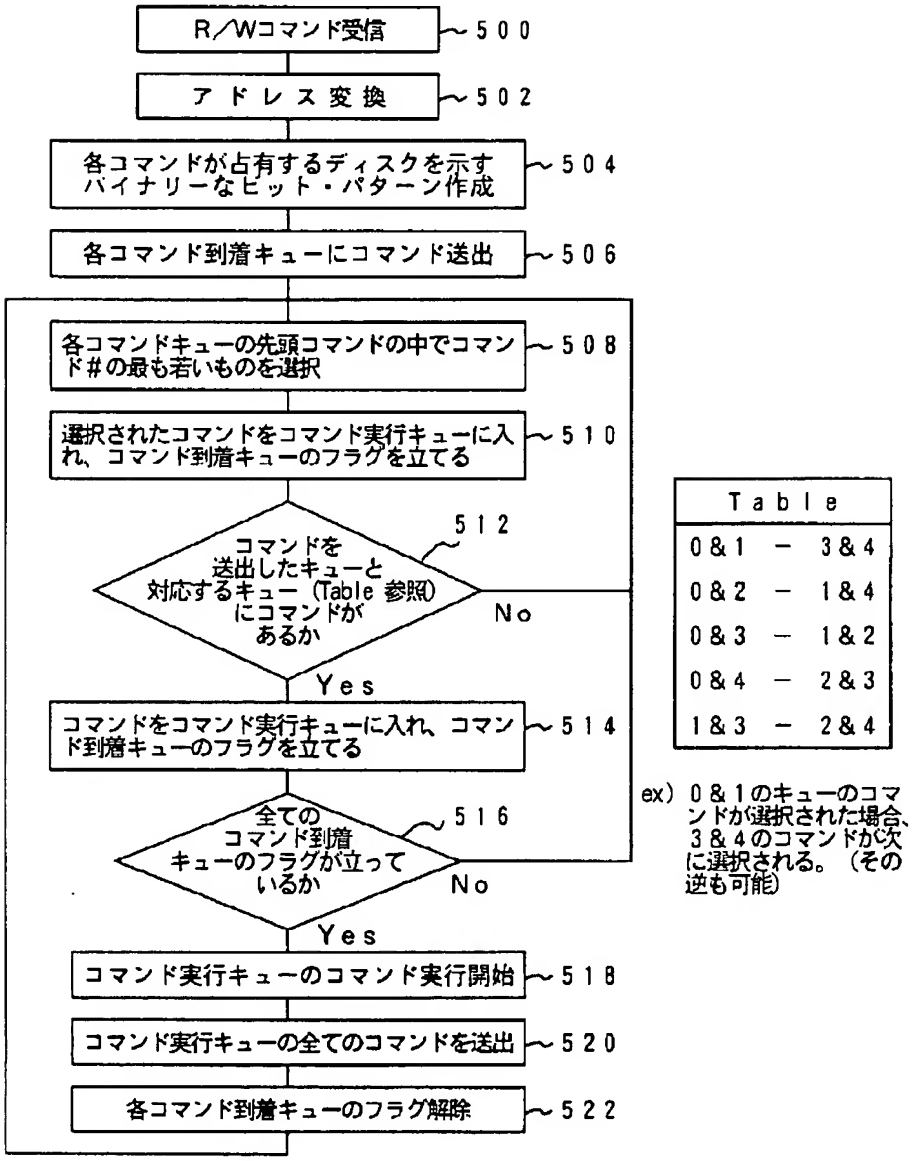
【図7】

図7　コマンド・スケジューラの第三実施例を示すシステムブロック図



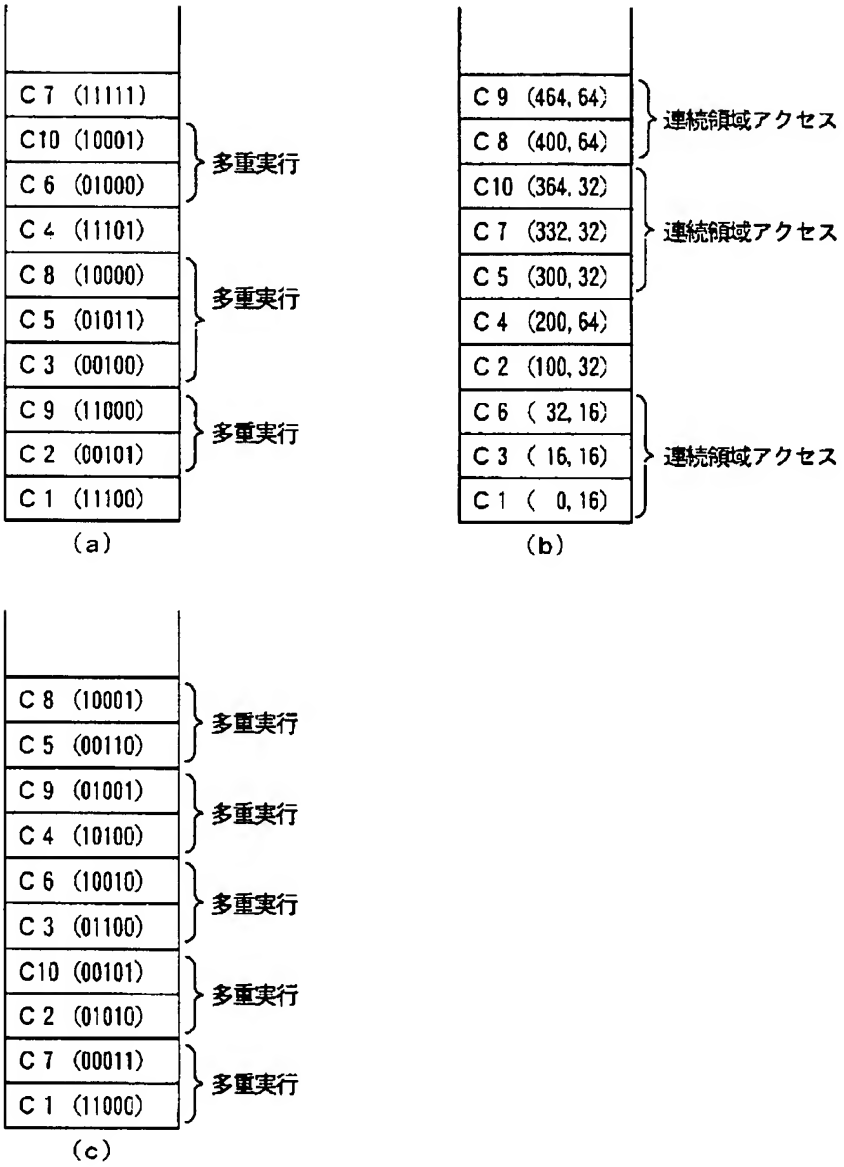
【図8】

図8　第三実施例の全体動作を示すフローチャート



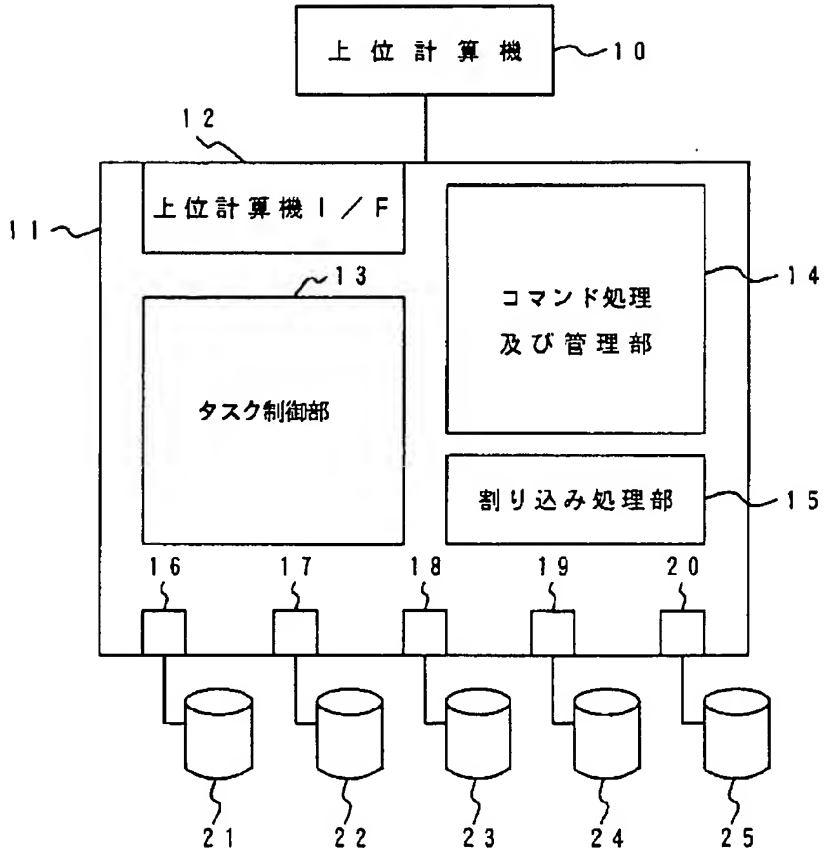
【図9】

図9　各実施例によるコマンドの並び変え結果を示す図



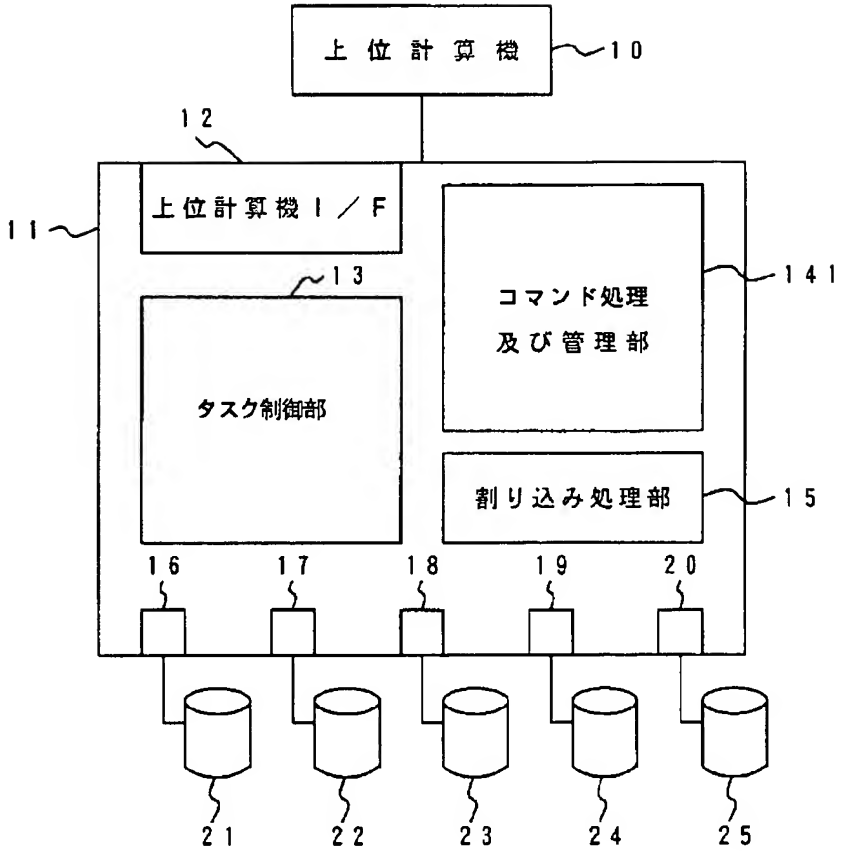
【図10】

図10


上位計算機
10
12
上位計算機I／F
11
13
タスク制御部
コマンド処理
及び管理部
14
割り込み処理部
15
16
17
18
19
20
21
22
23
24
25

【図11】

図11



フロントページの続き

(72)発明者　大枝　高
神奈川県横浜市戸塚区吉田町292番地　株式会社日立製作所マイクロエレクトロニクス機器開発研究所内

(72)発明者　松並　直人
神奈川県横浜市戸塚区吉田町292番地　株式会社日立製作所マイクロエレクトロニクス機器開発研究所内